P1 820696

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

June 21, 2002

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/287,946*
FILING DATE: *May 01, 2001*
RELATED PCT APPLICATION NUMBER: *PCT/US02/14141*

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

M. Tarver

M. TARVER
Certifying Officer

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

Appendix B

05-02-01 A/PRO

Please type a plus sign (+) inside this box [+]

Docket No. 5976

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

INVENTOR(S)		
Given Name (first and middle (if any))	Family Name or Surname	Residence (City and either State or Foreign Country)
Qing	Yang	Wakefield, Rhode Island

___ Additional inventors are being named on the ___ separately numbered sheets attached hereto.

TITLE OF THE INVENTION (280 characters max)

Distributed Web Server

CORRESPONDENCE ADDRESS

Direct all correspondence to:

___ Customer Number _____ (Type Customer Number here)

Place Customer Number Bar Code Label here

OR

_X_ Firm *or* Individual Name	Richard L. Stevens Samuels, Gauthier & Stevens, LLP				
Address	225 Franklin Street, Suite 3300				
City, State & ZIP	Boston, Massachusetts 02110				
Country	U.S.	Tel.	(617) 426-9180	Fax	(617) 426-2275

ENCLOSED APPLICATION PARTS *(check all that apply)*

_X_ Specification/Drawings *Number of Pages* 2

___ Other (specify) _____

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT *(check one)*

_X_ A check or money order is enclosed to cover the filing fees.

_X_ Applicant Claims Small Entity Status — FILING FEE AMOUNT

___ No fee is to be paid at this time.

_X_ The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Order Account Number: 19-0079 — $ 75.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

___ No.

_X_ Yes, the name of the U.S. Government agency and the Government contract number are: National Science Foundation, Grant No. CCR-0373377.

CERTIFICATION UNDER 37 C.F.R 1.10

I hereby certify that this correspondence and the documents referred to as attached therein are being deposited with the United States Postal Service on May 1, 2001 in an envelope as "EXPRESS MAIL POST OFFICE TO ADDRESSEE" service under 37 C.F.R. 1.10, Mailing Label Number EL821866557US addressed to the: Assistant Commissioner for Patents, Washington, D.C. 20231.

Julie A. Catalano
Signature

Julie A. Catalano
Type or print name of person certifying

Respectfully submitted,

Signature: _____ Date: 1 May 01

Typed or Printed Name: Richard L. Stevens
Registration No.: 24,445
Telephone: (617) 426-9180
Extension: 122

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

Boosting Web Server Performance Using DRALIC – Work in Progress

Xubin He, Qing Yang, Jian Li, and Ming Zhang
Dept. of Electrical and Computer Engineering
University of Rhode Island
Kingston, RI 02881
{hexb, qyang, lijian, mingz}@ele.uri.edu

I. Introduction

The goal of this work is to design and evaluate a new architecture called DRALIC—Distributed RAID And Location Independence Caching. DRALIC provides a direct and immediate solution to boost web server performance by making use of commodity computers that are available today. DRALIC starts working only when an actual disk request has come to the device no matter whether it is a result of file system miss or it is a request from a database operation. It does not require any change of existing operating systems, databases, nor applications. In one implementation, DRALIC combines the functions of disk I/O host bus adapter card (HBA) and the functions of the network interface card (NIC) to form an integrated I/O-Network card with a highly intelligent embedded-processor. Or in another implementation, DRALIC bridges the HBA and NIC by designing intelligent device drivers. Besides network accesses, the new interface card or drivers at each node control the local disk as well as a raw RAM partition of the system RAM of the node. The disk together with the ones in other nodes in the network forms a distributed RAID that appears to users as a large and reliable logic disk space. The raw RAM partitions in all nodes together form a large, global, and location independence cache for the RAID and is accessible to any node connected to the network, independent of its physical location. Therefore, DRALIC works at device or device driver level to allow all the nodes to work together in parallel to process web requests. The distributed RAID allows parallel operations of disk accesses and provides fault tolerance using parity disks, whereas location independence caches provide cooperative caching to the computing nodes for better I/O performance. Furthermore, DRALIC is a cost-effective architectural approach because it uses low cost PCs/Workstations that are often readily available as existing computing facilities in an organization or cooperation.

II. DRALIC Architecture

The main idea of DRALIC is very simple. It combines or bridges disk I/O host bus adapter card (HBA) and network interface card (NIC) to implement distributed RAID and global caching. Figure 1 shows the conceptual diagram of a DRALIC. A disk that exists in a PC/Workstation (node) is partitioned into two parts: one local disk that holds OS and local data and applications, and the other called DRALIC disk that is used by DRALIC. DRALIC disks in all nodes in the system are interconnected through the DRALIC controller and a network switch to form a distributed RAID. The system RAM in each node is also partitioned into two parts: one is controlled by local OS and the other, referred to as DRALIC RAM, is controlled by the DRALIC driver. The collection of DRALIC RAM in all nodes forms a unified system cache for the underlying RAID system.

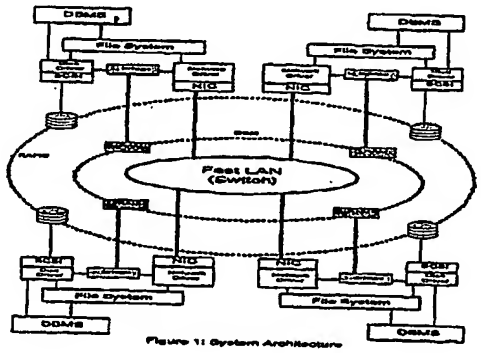

Figure 1: System Architecture

III. Preliminary Performance Analysis

To demonstrate the feasibility and performance potential of the proposed DRALIC, we present a preliminary performance analysis to look at the effects of bus and network delays on the performance potential of the DRALIC architecture. While our research will focus on System I/O, the current PCI bus can run at 33-132 MHz with data width of 32 or 64 bits. The memory bandwidth of PCI based system is BW_{mem}=33M*32bits/sec=132MB/sec. A Gigabit Ethernet switch with the transfer speed up to 1Gbps can provide network bandwidth approximately: BW_{net}=100MB/s. The overhead of network operation

including both software and hardware is assumed to be OH_{net}=0.2ms. As for disks, we consider a typical SCSI disk drive with specifications as the follows:

Model	Capacity	Average Seek Time	Rotational Speed	Average Latency	Transfer rate
UltraStar 18ES	9.1GB	7ms	7200RPM	4.17ms	187.2-243.7Mbps

Based on the above disk parameters, we can assume that a typical bandwidth of disk to be BW_{dsk}=25MB/s and the overhead of disk to be OH_{dsk}=12ms. The following is a list of notations and formulae used in our analysis:

B: data block size (8KB);
N: number of nodes within the DRALIC system;
H_{lm}: Local memory hit ratio;
H_{rm}: Remote memory hit ratio;
T_{lm}: Local memory access time (ms);
T_{rm}: Remote memory access time (ms);
T_{raid}: access time from the distributed RAID (ms);
T_{pc}: Average I/O response time of traditional PCs with no cooperative caching(ms);
T_{dralic}: Average I/O response time of DRALIC system (ms);

$$T_{lm}=\frac{1000B}{BW_{lm}}$$

$$T_{rm}=1000\left(\frac{B}{BW_{net}}+OH_{net}+\frac{B}{BW_{dsk}}\right)$$

$$T_{raid}=1000\left(\frac{(N-1)B}{N\times BW_{net}}+N\times OH_{net}+\frac{B}{N\times BW_{dsk}}+OH_{dsk}\right)$$

$$T_{pc}=1000\left(OH_{dsk}+\frac{B}{BW_{dsk}}\right)$$

$$T_{dralic}= H_{lm}\times T_{lm}+(1-H_{lm})\times H_{rm}\times T_{rm}+(1-H_{lm})\times(1-H_{rm})\times T_{raid}$$

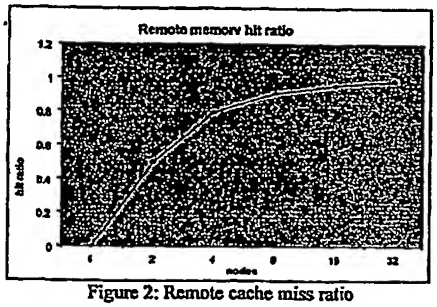

Figure 2: Remote cache miss ratio

With lack of measured hit ratios of remote caches, we assume remote hit ratio to be a logarithm function of number of nodes in the system as shown in Figure 2.

It is reasonable to assume that the remote cache hit ratio increases with the number of nodes because more nodes give larger cooperative cache spaces. The exact hit ratio number is not significant here since we use the hit ratio as a changing parameter to observe I/O performance as a function of it. From Figure 3, we can see that even with hit ratio of 50%, performance is doubled. With remote hit ratio of 80%, a factor of 4 performance improvement can be obtained. The data in this figure are sufficient to show the potential benefits of DRALIC.

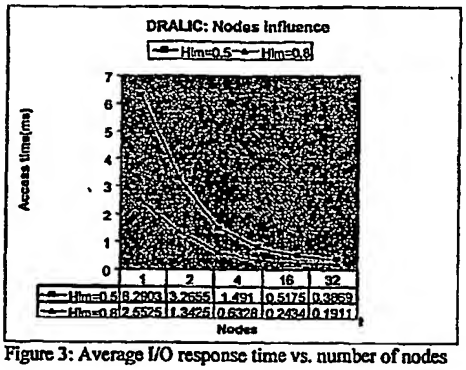

Figure 3: Average I/O response time vs. number of nodes

References

[1] T. E. Anderson, M. Dahlin, J. M. Neefe, D. A. Patterson, D. S. Roselli, R. Y. Wang. "Serverless Network File Systems." In *Proceedings of the Fifteenth ACM Symposium on Operating System Principles*, pp.109-126, Dec. 3-6, 1995

[2] G. A. Gibson, D. Nagle K. Amiri, J.Butler, F.W.Chang, H. Gobioff, C. Hardin, E. Riedel,D. Rochberg, J. Zelenka. "A Cost-Effective, High-Bandwidth Storage Architecture." *Proceedings of the 8th International Conference on Architectural Support for Programming Languages and Operating Systems(ASPLOS VIII)*, October 4-7,1998

[3] Y. Hu, Q. Yang, T. Nightingale. "RAPID-Cache: A Reliable and Inexpensive Write Cache for Disk I/O Systems." In *Proceedings of the 5th International Symposium on High Performance Computer Architecture* (HPCA-5), pp.204-213, Orlando, Florida. Jan. 9-13,1999

[4] Intel Developer Forum, *Peer to Peer Computing*, http://developer.intel.com/design/idf/fall2000/presentations/ptp/index.htm, Oct. 2000